

ML2020SPRING HW3 Report

學號：R08946015 系級：資料科學碩一 姓名：陳鈞廷

1. 請從 Network Pruning/Quantization/Knowledge Distillation/Low Rank

Approximation/Design Architecture 選擇兩者實做並詳述你的方法，將同一個大 model 壓縮至接近相同的參數量，並紀錄其 accuracy。

Design Architecture: 這次我參考了 torchvision 的 **MobileNetV2**，為了減少參數量，主要將 Inverted Residual 的數量從預設減少至 6 個，並減少每層 Inverted Residual 的 output channel 數量，6 層 Inverted Residual 的 out_channels 數量依序為: 16, 32, 32, 64, 64, 128，同時減少最後一層 Inverted Residual 的 expand ratio 至 1。最後總參數量減少至 **152,075**，所有參數都是 trainable。

Knowledge Distillation: 我參考了助教提供的 sample code 來實作，student net 使用上述的架構，teacher net 使用助教提供的 **ResNet18 ImageNet pretrained & fine-tune**。

	Validation Accuracy	# Parameters
Design Architecture	82.32%	152,075
Design Architecture + Know Distillation	79.88%	152,075

2. 請嘗試比較以下 accuracy (兩個 Teacher Net 由助教提供)以及 student 的總參數量以及架構，並嘗試解釋為甚麼有這樣的結果。你的 Student Net 的參數量必須要小於 Teacher Net 的參數量。(2%)

x. Teacher net architecture and # of parameters: torchvision's ResNet18, with 11,182,155 parameters.

y. Student net architecture and # of parameters:

Based on **MobileNetV2**，總參數量 **152,075**，所有參數都是 trainable。

a. Teacher net (ResNet18) from scratch: 80.09%

b. Teacher net (ResNet18) ImageNet pretrained & fine-tune: 88.41%

c. Your student net from scratch: validation acc. 最高達到 79.88%

d. Your student net KD from (a.): validation acc. 最高達到 80.81%

e. Your student net KD from (b.): validation acc. 最高達到 82.36%

從上面的 student Net 數據可以發現 performance 最好的是 KD from **ResNet18 ImageNet pretrained & fine-tune**，而最差的是沒有使用 KD 的模型。而在 from scratch 上 teacher net 和 student net 差距非常小，因此可以猜測 student net 已經有足夠的能力可以自行學習到跟 teacher net 在這個 food-11 task 學習到大部分的 feature。而 **ResNet18 ImageNet pretrained & fine-tune** 的 accuracy 跟其他還有一段蠻大的差距，因此從 ImageNet 學習的特徵抽取能力是這個 teacher net 的關鍵，這樣的猜測可以驗證在 d. 和 e. 的 accuracy 差距上，因此我認為 e. 有從 teacher net 上學到更多的 feature 抽取能力。

4. 請嘗試比較以下 validation accuracy，並且模型大小要接近 1MB: (2%)

- a. 原始 CNN model (用一般的 Convolution Layer) 的 accuracy
- b. 將 CNN model 的 Convolution Layer 換成總參數量接近的 Depthwise & Pointwise 後的 accuracy
- c. 將 CNN model 的 Convolution Layer 換成總參數量接近的 Group Convolution Layer (Group 數量自訂，但不要設為 1 或 in_filters)

為了節省訓練時間，以下三個 CNN model 皆訓練 50 個 epoch，並挑選訓練過程中最高的 validation accuracy。

	參數量	總 Conv2d 層數	Validation Acc.
原始 CNN	264,075	4	68.425%
Depthwise & Pointwise	270,639	4 (DW) + 4(PW)	71.428%
Group Convolution	258,955	5	64.839%